AD_____

Award Number: DAMD17-00-1-0048

TITLE: Structure/Function Studies of the Androgen Receptor DNA-Binding Region

PRINCIPAL INVESTIGATOR: Fraydoon Rastinejad, Ph.D.

CONTRACTING ORGANIZATION: University of Virginia
Charlottesville, Virginia 22906

REPORT DATE: April 2001

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	April 2001	Annual (1 Apr 00 - 31 Mar 01)

4. TITLE AND SUBTITLE
Structure/Function Studies of the Androgen Receptor DNA-Binding Region

5. FUNDING NUMBERS
DAMD17-00-1-0048

6. AUTHOR(S)
Fraydoon Rastinejad, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Virginia
Charlottesville, Virginia 22906

E-Mail : fr9c@virginia.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 Words)*

The onset and progress of prostate cancer is related to the function of the androgen receptor (AR), and much of current drug therapy is directed against this protein. AR contains two binding sites: a ligand binding domain (LBD) and a DNA binding domain (DBD) through which AR regulates gene expression. The proposal focuses on the AR DBD and its complex with androgen response elements. Our overall goals are to identify the appropriate protein and DNA constructs that will allow us to co-crystallize relevant DBD-DNA complexes for structure determination. We have co-crystallized a minimal AR DBD with a 19 base-pair sequence representing a consensus response site. Because our previous work on nuclear receptor - DNA complexes has shown that the linker region between the DBD and LBD provides a significant DNA interface, we are also characterizing new AR DBDs in which the DBD is extended at its C-terminal region by various sized linker regions. These constructs are being tested for binding to DNA sequences representing various classes of androgen response elements. Finally, to understand how other types of response elements are recognized, we are testing a series of DNA duplexes for protein binding to use in upcoming co-crystallization trials.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Androgen Receptor, X-ray Crystallography, DNA-protein interactions	10
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

Structure/Function Studies of Androgen Receptor DNA Binding Region

Introduction

The Androgen Receptor:
Androgens (testosterone, dihydrotestosterone) are steroids with a key role in promoting normal sex differentiation and development, pubertal masculinization, initiation of spermatogenesis, and maintenance of male sexual function [1-3]. There is evidence for androgen production in the testes, skin, and submaxillary glands, although its levels are most abundant in the prostate gland. The role of androgens in the control of normal and abnormal sexual development in humans and other vertebrates has been studied since 1935- when the chemical structure of testosterone was first elucidated [1]. The androgen receptor (AR) is a ligand-activated transcription factor and a member of a large superfamily of related receptors [4-7].

All the regulatory actions produced by androgens occur at the level of transcription initiation in which a hierarchy of intermolecular interactions are based on the DNA-bound complex of the androgen receptor (AR) [4-9]. The AR gene contains an open reading frame separated over eight exons, producing a protein of 919 amino-acids in humans with four discrete functional domains, shown in Figure 2 [2,3,10]. The DNA-binding domain (DBD) forms into a homodimer, but only on the specific DNA binding sites of AR. A similar DNA-dependent dimerization mode is used by other steroid receptors and many other members of the nuclear receptor superfamily, although for non-steroid receptors the DBDs can also form heterodimers with the 9-cis retinoic acid receptor (RXR) [6].

The core DBD of nuclear receptors are highly conserved across the nuclear receptor family (typically 50-60% identical in amino-acid sequence). Because this region is so conserved, the origins of DNA-binding specificity have remained somewhat elusive until the recent structure determination of related receptor DNA-binding complexes. Our own laboratory's crystal structures of related DBDs have shown a special role for the hinge region, which imparts selectivity in DNA minor groove recognition and also provides the unique DNA-dependent dimerization in some cases [11-14]. This region is not conserved in size or sequence across the superfamily, further underscoring its unique role in specifying the DNA-binding site for each particular receptor.

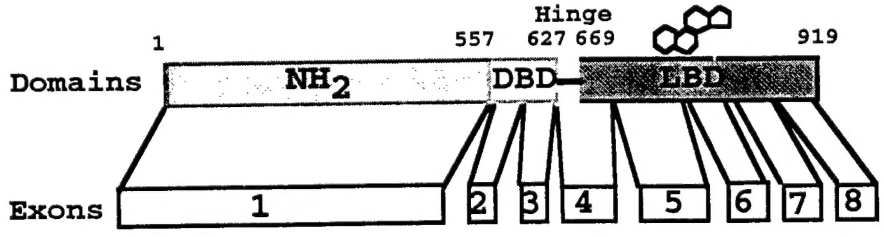

Figure 1: The Domain and Exon representation of the human Androgen Receptor. Two independent and separable functions of the nuclear receptors are: the DNA-binding domain (DBD) confers response element recognition, and the Ligand-Binding domain (LBD) binds the steroid hormone (shown schematically).

The Androgen Response Elements:

All the steps in androgen signaling, occur only through the DNA-bound complex of this receptor [2,4,5]. The specific DNA-binding sequences in the genome are known as response elements. A number of such genes have been shown to contain AR DNA response elements in upstream regulatory sites. The AR response element DNA is a nearly palindromic sequences, which, as expected, accommodates the homodimeric form of AR DBDs, and is superficially related to the response sites of the glucocorticoid and estrogen receptors. In these response sites, two consensus hexameric stretches of the sequence 5'AGAACA-3' are separated by three base pairs [2,4,5]. However, naturally occurring androgen response sites usually contain significant deviations from the consensus half-site sequences, and also contain unique sequences in the flanking regions which increase the binding affinity and allow target site selectivity amongst the related steroid receptors [15]. Such androgen response elements have been identified upstream of the probasin gene and also the prostate specific antigen (PSA, which is also used as the diagnostic marker for prostate cancer [16].

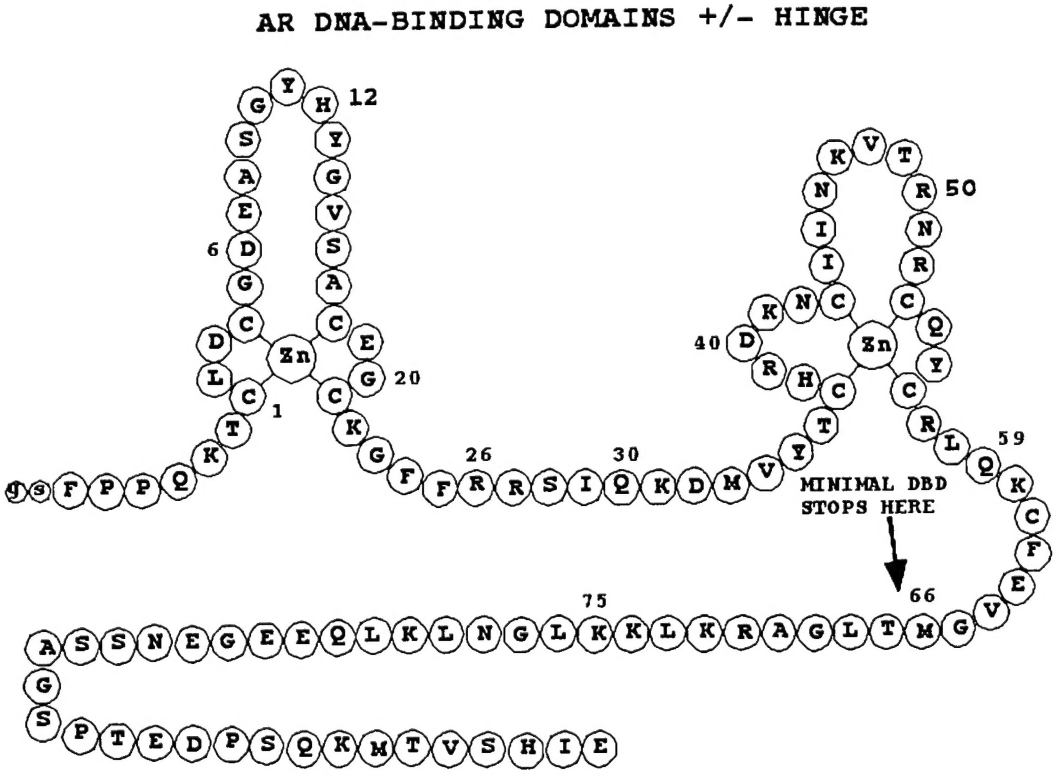

Figure 2. The AR core DBD containing a C-terminal extention into the receptor's hinge region. Residues numbered from 1 to 66 represent the core DBD, and residues beyond 66 are from the hinge region of the receptor (shown schematically in Figure 1).

Goals and Key Research Accomplishments over the previous Year

One of our two primary aims has been to work toward crystallization and X-ray crystal structure determination of the minimal androgen receptor (AR) DNA-binding Domain (DBD) in complex with a 19 base pair idealized DNA target. It is important to note that the DNA target used here was quite similar to the binding site generally attributed for other steroid receptors such as the glucocorticoid and mineralcorticoid receptors, and thus is likely not to provide all the information about the target selectivity of AR. For this reason, we have a second goal (described below) that will address this specific question more thoroughly. Nevertheless, we have been able to grow these crystals (Figure 3) and improve them in terms of their size and quality. We have tested several cryogen protectants for use in synchrotron data collection under cryogenic temperatures (where we are likely to achieve the best diffraction), and have also scheduled a synchroton trip for later this year for collecting a complete diffraction dataset. To assist our ability to solve the structures, we have also devised a number of useful search models for use by molecular replacement. These search models consist of common sequences shared by other DBDs whose structures we have previously solved in the laboratory. Other, AR specific amino-acids are trimmed to alanines to assist in the molecular replacement search. We have tested this strategy on other DBD/DNA complexes involving nuclear receptors and believe that it is likely to prove successful for solving the structure of our minimal AR DBD/DNA complex.

Figure 3: Crystals of our AR DBD complex with a 19 mer DNA duplex. The DNA is an idealized steroid response element composed of two AGAACA sequences arranged symmetrically about a three base-pair spacer. The DBD is a minimal sequence composed mainly of the AR core sequence.

The second aim of the proposal is to pursue crystals of other DNA binding complexes of AR which are more informative in terms of binding selectivity of this receptor with respect to glucocorticoid, mineralcorticoid, and other steroid receptors. We have taken a two tiered approach in our initial biochemical studies towards identifying the most useful candidates for structural examination. First, we have made a series of AR DNA binding regions in which the core DBD is extended at its C-terminal end with various sized hinge region segments in order to map out more precisely the boundaries of the protein to be used in crystallization. Based on previous work we and others have carried out with nuclear receptor DBDs, we strongly believe that additional sequences beyond the core AR DBD are likely to have major consequences in terms of target DNA selectivity and

cooperative homodimerization. Figure 3 shows the extension of the core DBD in the case of the androgen receptor (residues beyond 66), and Figure 4a shows the specific constructs that we made for over-expression in *E.coli*. The five constructs that we now have all share a common N-terminal $(His)_6$ tag for purification, the 67 residue core DBD region, plus 0, 10, 20, 30, 41 residues of hinge region residues at their C-terminus, respectively. The constructs were generated by PCR and cloned into pET16b vector, and expressed in BL21-DE3 strain. A Ni-NTA column and an S column are used in succession to purify the proteins (all of which proved to be in the soluble fraction of *E.coli*). Figures 4b-c shows an example of how each of these constructs can be purified to homogeneity. All of this work represents new achievements in the past twelve months, and will significantly guide us in generating the most important and useful co-crystal structures in the upcoming two years for meeting our overall goals.

```
Construct 1: MGHHHHHHIPQKTCLICGDEASGCHYGALTCGSCKVFFKRAAEGKQKYLCASRNDCT
Construct 2: MGHHHHHHIPQKTCLICGDEASGCHYGALTCGSCKVFFKRAAEGKQKYLCASRNDCT
Construct 3: MGHHHHHHIPQKTCLICGDEASGCHYGALTCGSCKVFFKRAAEGKQKYLCASRNDCT
Construct 4: MGHHHHHHIPQKTCLICGDEASGCHYGALTCGSCKVFFKRAAEGKQKYLCASRNDCT
Construct 5: MGHHHHHHIPQKTCLICGDEASGCHYGALTCGSCKVFFKRAAEGKQKYLCASRNDCT

Construct 1: IDKTRRNNCPSCRLRKCYEAGM
Construct 2: IDKTRRKNCPSCRLRKCYEAGMTLGARKLKKL
Construct 3: IDKTRRKNCPSCRLRKCYEAGMTLGARKLKKLGNLKLQEEGE
Construct 4: IDKTRRKNCPSCRLRKCYEAGMTLGARKLKKLGNLKLQEEGENSSAGSPTED
Construct 5: IDKTRRKNCPSCRLRKCYEAGMTLGARKLKKLGNLKLQEEGENSSAGSPTEDPSQKMTVSHIE
                                   +10       +20       +30       +41
```

Figure 4a: Five AR DNA binding constructs that were made with varying length of hinge region. We are testing each of these using electrophoretic mobility shift assays to characterize their binding to various AR response elements. The MGHHHHHH sequence at the beginning of each construct represents a' his-tag sequence' that is added for ease of protein purification. The methionines residue at the end of construct 1represents residue 66 in Figure 2, and constructs 2-5 contain an additional 10-40 residues from the AR hinge region.

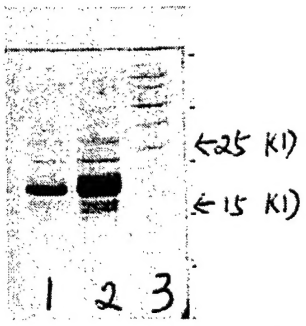

Fig 4. Construct 5 eluted from Ni-NTA column Lane 1, 2: fractions eluted from Ni-NTA column. Lane 3: Molecular weight marker

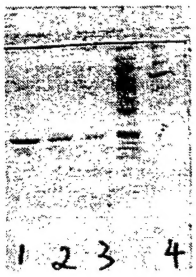

Fig 4. Construct 5 eluted from S column. Lane 1, 2, 3: fractions eluted from S column; Lane 4: Molecular weight marker.

The second approach we have taken is to find the best candidates for structural studies is to examine androgen-responsive elements which represent highly selected and natural targets for AR. Unlike the DNA contained in crystals described in Aim 1, the naturally occurring androgen response elements in this aim contain a) significant differences from the consensus steroid element, b) additional flanking sequences outside of the core symmetric sequence, and c) less extensive two-fold symmetry. Therefore, a structural such an additional complex, in comparison with the structure we are pursuing in Aim 1, will likely teach us considerable new lessons about AR DNA binding specificity and selectivity.

A recent report (By Dr. Colleen Nelson and colleagues) characterized several different types of androgen response elements that occur in the promoter of an important androgen responsive gene, the probasin gene (see Figure 5). These response elements, together with those identified earlier in the PSA promoter, can all be categorized into basically two major classes. Class I sequences are more typical of conventional steroid response elements with the sequence RGAACA-NGN-TGTNCT. Class II AREs were only discovered by methylation protection assay in the presence of androgen receptor and do not share all the hallmarks of Class I. The class II consensus sequence is RGGACA-NNA-AGCCAA. It has been suggested that appropriate combination of class I and class II AREs, as that happened *in vivo*, can lead to allosteric or perhaps differential binding. So we expend to expand our studies by using the five protein constructs in Figure 4a and the two major classes of naturally occurring AREs in Figure 6, to identify the most functionallhy revealing constructs for our upcoming c-cyrstallization trials.

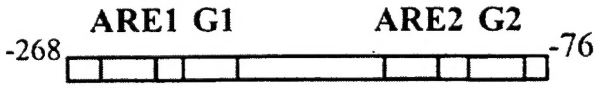

Fig 5. Probasin Promoter Structure. There are 4 different AREs: ARE1, G1, ARE2 and G2 over the –268 to –76 region of the promoter.

Probasin G-1	-209GGGACA-TAA-AGCCCA-196
Probasin G-2	-107ATGACA-CAA-TGTCAA-93
PSA Enhancer V	-4234 GGGACA-ACT-TGCAAA-4220
PSA Enhancer IIIA	-4079 AGGACA-GTA-AGCAAG-4065
PSA Enhancer IV	-4175 AGATCA-TGA-AGATAA-4189
SLP 2	+143*AGAACT-GGC-TGACCA*+128
	• ••••
CONSENSUS CLASS II	RGGACA-NNA-AGCCAA
	-7 -2 0 +2 +7
CONSENSUS CLASS I	RGAACA-NGN-TGTNCT
	o o o
Probasin ARE1	-241ATAGCA-TCT-TGTTCT-227
Probasin ARE2	-136AGTACT-CCA-AGAACC-122
PSA Enhancer III	-4148 GGAACA-TAT-TGTATT-4134
PSA ARR	-390*GGATCA-GGG-AGTCTC*-376
PSA ARE	-167*AGAACA-GCA-AGTGCT*-153
SLP 3	+144*AGAACA-GGC-TGTTTC*+158

Fig 6. Class I and Class II AREs that we are pursuing in our studies.

To assay protein-DNA binding, we have been relying on a sensitive electrophoresis mobility shift assay (EMSA). So far, four different simple AREs from probasin promoter, including two class I (ARE1, ARE2), and two class II (G-1, G-2) have been synthesized,

in each case having one strand fluorescein end-labeled. For comparison, the highest affinity but unnatural simple ARE GGTACAnnnTGTTCT has also synthesized and end-labeled. We have carried out several EMSA assays and are working to complete our studies. In each case, to determine their relative binding affinity, constant amount of DNA plus increasing amount of protein is used (see Figure 7 for example).

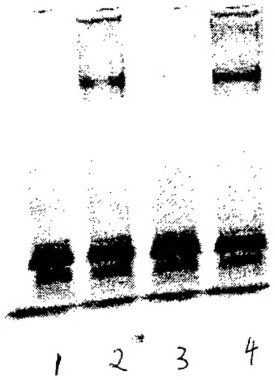

Figure 7. EMSA experiment characterizing the efficiency of AR construct 5 (figure 4) binding to a type II ARE. Lane 1: Free DNA (ARE2; 40nM); lane 2: DNA plus 1:15 (mol ratio) protein; lane 3: DNA plus 1:15 protein plus 2uM poly dIdC; lane 4: DNA plus 1:30 protein. Noted the binding affinity is very low here, may need to optimize binding condition.

Reportable Outcomes

Five new over-expression constructs for producing the AR DNA binding region.

Conclusions

We are in a position to determine important co-crystal structures of the Androgen receptor DNA binding complex with one or more response elements. Once we identify the most important crystallization targets and determine their structure, we will be well positioned to understand a) structure/function relationships in terms of the AR DNA binding region and androgen response elements, b) the basis for certain androgen insensitivity syndrome mutations that fall in the DNA-binding domain, and c) the possibility of whether the protein-DNA complex can be viewed as a useful drug target.

Our goal for the next twelve months is to complete our biochemical studies on protein and DNA constructs, identify the high affinity complexes that are likely to yield additional useful crystals, and work toward the structure determination. If our EMSA assays indicate that various different response elements show high affinity binding to AR DNA binding regions, we are likely to pursue all of these for structure determination to understand how class I, class II and the consensus ARE differ in their protein-DNA contacts.

References

1. Endorinology. 4th ed. Hadley, ME. 1997. Prentice hall (NJ).
2. Androgen receptor: an overview. Chang C. Saltzman A. Yeh S. Young W. Keller E. Lee HJ. Wang C. Mizokami A. Critical Reviews in Eukaryotic Gene Expression. 5(2):97-125, 1995.
3. Androgen action. Roy AK. Chatterjee B. Critical Reviews in Eukaryotic Gene Expression. 5(2):157-76, 1995.
4. Steroid hormone receptors: many actors in search of a plot. Beato M. Herrlich P. Schutz G. Cell. 83(6):851-7, 1995.
5. The nuclear receptor superfamily: the second decade. Mangelsdorf DJ. Thummel C. Beato M. Herrlich P. Schutz G. Umesono K. Blumberg B. et al. Cell. 83(6):835-9, 1995 Dec 15.
6. The RXR heterodimers and orphan receptors. Mangelsdorf DJ. Evans RM. Cell. 83(6):841-50, 1995.
7. Molecular mechanisms of action of steroid/thyroid receptor superfamily members. Tsai MJ. O'Malley BW. Annual Review of Biochemistry. 63:451-86, 1994.
8. The presence of a transcription activation function in the hormone-binding domain of androgen receptor is revealed by studies in yeast cells. Moilanen A. Rouleau N. Ikonen T. Palvimo JJ. Janne OA. FEBS Letters. 412(2):355-8, 1997.
9. Delineation of two distinct type 1 activation functions in the androgen receptor amino-terminal domain. Chamberlain NL. Whitacre
10. Localization of functional domains in the androgen receptor. MacLean HE. Warne GL. Zajac JD. Journal of Steroid Biochemistry & Molecular Biology. 62(4):233-42,
11. Structural determinants of nuclear receptor assembly on DNA direct repeats [see comments]. Rastinejad F. Perlmann T. Evans RM. Sigler Nature 375(6528):203-11, 1995.
12. Zhao, Q., Khorasanizadeh S, miyoshi, Y. Lazar M., Rastinejad F. Structural elements of an orphan nuclear receptor-DNA complex. Molecular Cell 1: 849-61, 1998.
13. Zhao, Q., Chasse, S., Devarakonda, S., Sierk, M.L. Sigler, P.B., and Rastinejad, F. Structure and Mechanism of RXR homodimer formation on DNA direct repeats. *Nature Structural Biology*. (*In press*).
14. Wagner, T., Khorasanizadeh, S., Chasse, S., Zhao, Q., and Rastinejad, F. Structure of the ternary DNA-binding complex of RXR-RAR on a negative regulatory DNA site (manuscript in preparation).
15. Characterization of Two Cis-Acting DNA Elements Involved in the Androgen Regulation of the Probasin Gene. Rennie, et al., Mol. Endocrinology, 7:23-36,1993.
16. The androgen receptor in prostate cancer. . J. Brinkmann AO. Pathology, Research & Practice. 192(7):752-60, 1996.